1 Problem Statement

Gaussian linear models are often insufficient in practical applications, where noise can be heavy-tailed. In this problem, we consider a linear model of the form $y_i = a \cdot x_i + b + e_i$. The (e_i) are independent noise from a distribution that depends on x as well as on global parameters; however, the noise distribution has conditional mean zero given x. The goal is to derive a good estimator for the parameters a and b based on a sample of observed (x, y) pairs.

1.1 Instructions:

1. Load the data, which is provided as (x, y) pairs in CSV format. Each file contains a data set generated with different values of a and b. The noise distribution, conditional on x, is the same for all data sets.

2. Formulate a model for the data-generating process.

3. Based on your model, formulate a loss function for all parameters: a, b, and any additional parameters needed for your model.

4. Solve a suitable optimization problem, corresponding to your chosen loss function, to obtain point estimates for the model parameters.

5. Formulate and carry out an assessment of the quality of your parameter estimates.

6. Try additional models if necessary, repeating steps $2 - 5$.

1.2 Deliverables:

1. A CSV file containing your point estimates of the linear model parameters a and b for each of the input files. The output CSV file will have three columns: input file name, a, and b corresponding to your solution for that data set.

2. A brief report on your approach. Discuss modeling assumptions you made as well as any hand-picked numerical constants. Please submit the report in pdf format; include any necessary figures and equations.

3. Your code for deriving and assessing the estimates. We will evaluate your code for correctness, conciseness, and readability (comments and documentation). We prefer R or Python, but feel free to use any high-level programming language you are comfortable with. You are allowed to use any publicly available software packages.